MAF+ RECORD LINKAGES AND FRAMES ENHANCEMENTS

Decennial Census Programs - Geography Division

Michael Ratcliffe —Senior Advisor for Frames

Richard Watson —Geospatial Frame Team Lead

Jennifer Ortman, Ph.D. —Principal Demographer, Population Division

Lori Zehr, Ph.D. —Census Frames Technical Lead

CIVIC DIGITAL FELLOWSHIP

United States Census Bureau

AISHANI DUTTA

University of Illinois at Urbana-Champaign

Computer Science

EVAN DONG

Brown University

Computer Science

*The following slides contain NO Title 13 information

THE MAF ENHANCEMENT PROJECT

Goals:

- Make reliable, long reaching linkages between Census Data Frames
- Reduce duplication, manual efforts, and one off data errors
- Reuse existing data to reduce the burden on respondents
- Create a coordinated and integrated environment for disparate frames

CIVIC DIGITAL FELLOWSHIP

United States Census Bureau

GEOSPATIAL FRAME GOALS

Develop a system for parsing, standardizing, and matching addresses

Reduce the manual burden involved in deduplicating records, updating frames, identifying linkages, and finding errors

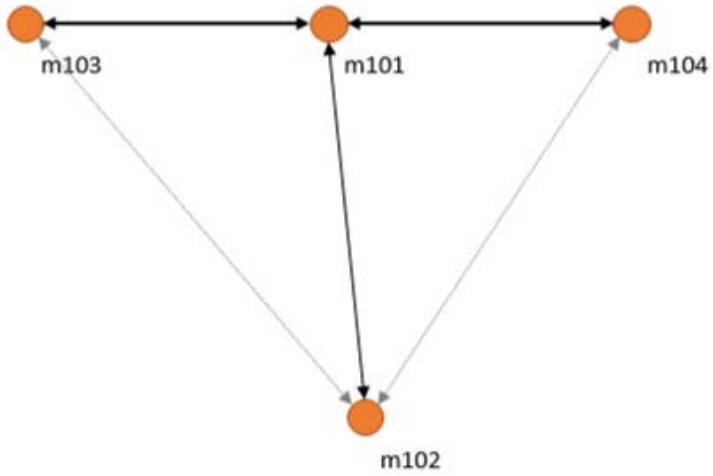
Guard against data discrepancies, reduce NRFU costs, expand the Bureau's ability to report on changes and trends

CIVIC DIGITAL FELLOWSHIP

United States Census Bureau

GEOSPATIAL FRAME PROGRESS

- **Standardizing Tools**
  - an improved version of the legacy standardizer built upon an open-source parser (usaddress)
- **Record Matching Tools**
  - optimized system to deduplicate records and identify preliminary matches across datasets
- **Preprocessing Tools**
  - number conversions, phonetic matches, spelling-check



Core Address	
MPCID	Address
m101	101 Main Street
m102	101 Main Ave
m103	101 N Main St
m104	101 S Main St

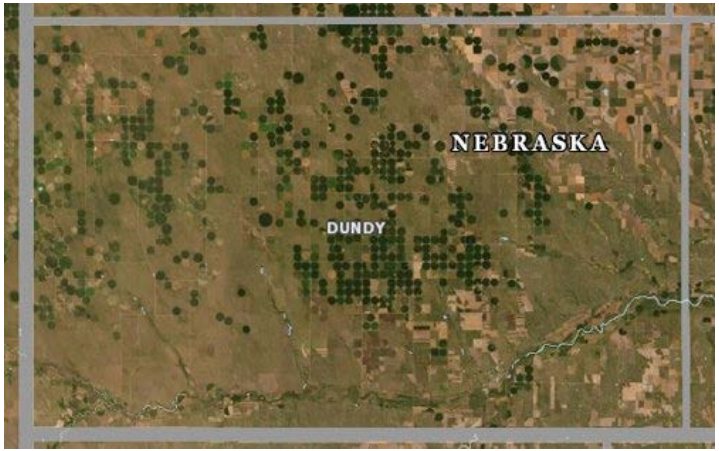
Note: This slide does not contain Title 13 data.

DEMOGRAPHIC FRAME GOALS

Create a single database to store data about individuals and their characteristics consolidated from several sources

Our prototype:

- Aggregate 2010 Census + American Community Survey (ACS) data
- Start with person -level data from one county
- Small-scale SQLite/serverless setup
- Focus on decennial census questions



Note: This slide does not contain Title 13 data.

DEMOGRAPHIC FRAME PROGRESS

Successes:

- Report on potential data platforms + infrastructure requirements
- SQLite schema and prototype for ACS data & Census Image Retrieval Application (CIRA) data

Lessons Learned:

- Acquiring server space and software
- Data acquisition
- Disparate file storage locations
- Data use agreements
- Iterative stage development and planning

CIVIC DIGITAL FELLOWSHIP

United States Census Bureau